

ABSTRACT

A document categorizing apparatus includes a sentence analyzer 12 for analyzing a plurality of documents to detect titles thereof; a feature element extractor 13 for extracting feature elements from the titles detected by the sentence analyzer 12 from the respective documents; feature table generating means 14 for generating a feature table representing the relationships between the feature elements extracted from the title and the documents including the feature elements; a document categorizing unit 15 for categorizing the documents into a plurality of clusters according to semantic similarity on the basis of the content of the feature table; a categorization result storage unit 16 for storing the clusters created by the document categorization unit 15; a cluster merging unit 2 for performing a cluster merging process upon the clusters stored in the categorization result storage unit 6; and an output control unit 31 for outputting the result of the cluster merging process to a display unit 32.

100762126-000004